

CLAIMS:

- Sub
a2
1. A method for classifying a text object, comprising:
- 5 extracting a set of features from the text object;
- constructing a document class fuzzy set with ones of the set of features
- 5 extracted from the text object; each of the ones of the features extracted from
- the text object having a degree of membership in the document class fuzzy
- set and a plurality of class fuzzy sets of a knowledge base;
- measuring a degree of match between each of the plurality of class
- fuzzy sets and the document class fuzzy set; and
- 10 using the measured degree of match to assign the text object a label
- that satisfies a selected decision making rule.
2. The method according to claim 1, further comprising learning each
- class fuzzy set in the knowledge base.
- 15 3. The method according to claim 2, wherein each class fuzzy set is
- learned by:
- obtaining a set of class training documents;
- merging those training documents in the set of training documents with
- similar labels to create a class document; and
- 20 computing a class fuzzy set using the class document.
4. The method according to claim 1, wherein the set of features is
- extracted from the text object by:
- tokenizing the document to generate a word list;
- parsing the word list to generate the set of grammar based features;
- 25 and
- filtering the set of grammar based features to reduce the number of
- features in the set of grammar based features to define the ones of the set of
- features extracted from the text object used to construct the document class
- fuzzy set.

5. The method according to claim 1, wherein the document fuzzy set is computed by:

calculating a frequency of occurrence for each feature in the set of features in the text object;

5 normalizing the frequency of occurrence of each feature in the set of features; and

transforming the normalized frequency of occurrence of each feature in the set of features to define the document fuzzy set.

6. The method according to claim 6, wherein the normalized frequency of occurrence of each feature in the set of features is transformed using a bijective transformation.

7. The method according to claim 1, wherein the degree of match between each of the plurality of class fuzzy sets and the document fuzzy set is measured using one of a maximum-minimum strategy and a probabilistic reasoning strategy based upon semantic unification.

8. The method according to claim 1, further comprising:

filtering each degree of match with an associated class specific filter function to define an activation value for its associated class rule;

identifying the activation value of the class rule with the highest activation value; each class rule having an associated class label; and

assigning the class label of the class rule with the highest identified activation value to classify the text object into one of the plurality of class fuzzy sets.

9. The method according to claim 8, further comprising learning each associated class specific filter function.

10. The method according to claim 1, wherein the decision making rule is used to identify one of a maximum value, a threshold value, and a predefined number.

11. A method for classifying a text object, comprising:

extracting a set of granule features from the text object;

constructing a document granule feature fuzzy set using ones of the granule features extracted from the text object; each of the ones of the granule features extracted from the text object having a degree of membership in a corresponding granule feature fuzzy set of the document granule feature fuzzy set and a plurality of class granule feature fuzzy sets of a knowledge base;

computing a degree of match between each of the plurality of class granule feature fuzzy sets and the document granule feature fuzzy set to provide a degree of match for each of the ones of the granule features;

aggregating each degree of match of the ones of the granule features to define an overall degree of match for each feature; and

using the overall degree of match for each feature to assign the text object a class label that satisfies a selected decision making rule.

12. The method according to claim 11, further comprising filtering the granule features extracted from the text object to define the ones of the granule features used to construct the document granule feature fuzzy set.

13. The method according to claim 12, wherein the filtering of the granule features is based upon one of Zipf's law and semantic discrimination analysis.

14. The method according to claim 11, wherein each granule feature is represented by a plurality of fuzzy sets and associated labels.

15. The method according to claim 11, wherein the ones of the granule features that are used to construct the document granule feature fuzzy set are reduced to one of a predefined threshold number of granule features and range of granule features.

16. The method according to claim 11, further comprising learning each granule fuzzy set in the knowledge base.

17. The method according to claim 11, wherein the degree of match between each of the plurality of class granule feature fuzzy sets and the document granule feature fuzzy set is measured using one of a maximum-minimum strategy and a probabilistic reasoning strategy based upon semantic unification.

18. The method according to claim 11, further comprising:

filtering each degree of match with an associated class specific filter function to define an activation value for its associated class rule;

identifying the activation value of the class rule with the highest activation value; each class rule having an associated class label; and

assigning the class label of the class rule with the highest identified activation value to classify the text object into one of the plurality of class granule feature fuzzy sets.

19. The method according to claim 18, further comprising learning each associated class specific filter function.

20. The method according to claim 11, wherein each degree of match is aggregated using one of a product and an additive model.

21. The method according to claim 11, further comprising estimating granule feature weights when they are aggregated as a weighted function.

22. A text categorizer for classifying a text object, comprising:

a knowledge base for storing categories represented by fuzzy sets and associated class labels;

a pre-processing module for representing extracted features from the text object as a document fuzzy set; and

an approximate reasoning module for using a measured degree of match between the fuzzy sets in the knowledge base and the document fuzzy set to assign the associated class labels of those categories that satisfy a selected decision making rule.

23. The text categorizer of claim 22, wherein the fuzzy sets are granule fuzzy sets.

24. The text categorizer of claim 22, wherein the fuzzy sets are class fuzzy sets.

5 25. The text categorizer of claim 24, further comprising a learning module for learning the class fuzzy sets.

26. The text categorizer of claim 25, further comprising:

a training database for creating a plurality of class documents; and

10 a validation database for validating learned class fuzzy sets in the knowledge base.

TELETYPE
A2